(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 1 074 974 A2

(12) EUROPEAN PATENT APPLICATION

(43) Date of publication:
07.02.2001 Bulletin 2001/06

(51) Int Cl.[7]: **G10L 17/00**

(21) Application number: **00304841.0**

(22) Date of filing: **07.06.2000**

(84) Designated Contracting States:
AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU MC NL PT SE
Designated Extension States:
AL LT LV MK RO SI

(30) Priority: **07.06.1999 US 326717**

(71) Applicant: **NOKIA MOBILE PHONES LTD.**
02150 Espoo (FI)

(72) Inventors:
- **Ulvinen, Vesa**
  90800 Oulu (FI)
- **Paloniemi, Jari**
  90900 Kiminki (FI)

(74) Representative: **Read, Matthew Charles et al**
Venner Shipley & Co.
20 Little Britain
London EC1A 7DH (GB)

(54) Secure wireless communication user identification by voice recognition

(57) A method to authorize or authenticate a user of a wireless telecommunication system (32), and includes steps of (a) selecting a word at random from a set of reference words, or synthesizing a reference word; (b) prompting the user to speak the reference word; and (c) authenticating the user to operate in or through or with a resource reachable through the wireless telecommunication system, only if the user's speech characteristics match pre-stored characteristics associated with the reference word. In one embodiment the steps of selecting or synthesizing, prompting and authenticating are performed in a mobile station (10) having a speech transducer (19) for inputting the user's speech, while in another embodiment at least one of the steps of selecting or synthesizing, prompting and authenticating are performed in a wireless telecommunications network (32) that is coupled between the mobile station and a telephone network (35). In yet another embodiment at least one of these steps are performed in a data communications network resource (38) that is coupled through a data communications network (37), such as the Internet, and the wireless telecommunications network to the mobile station. The step of prompting may include a step of displaying alphanumeric text and/or a graphical image to the user using a display (20) of the mobile station.

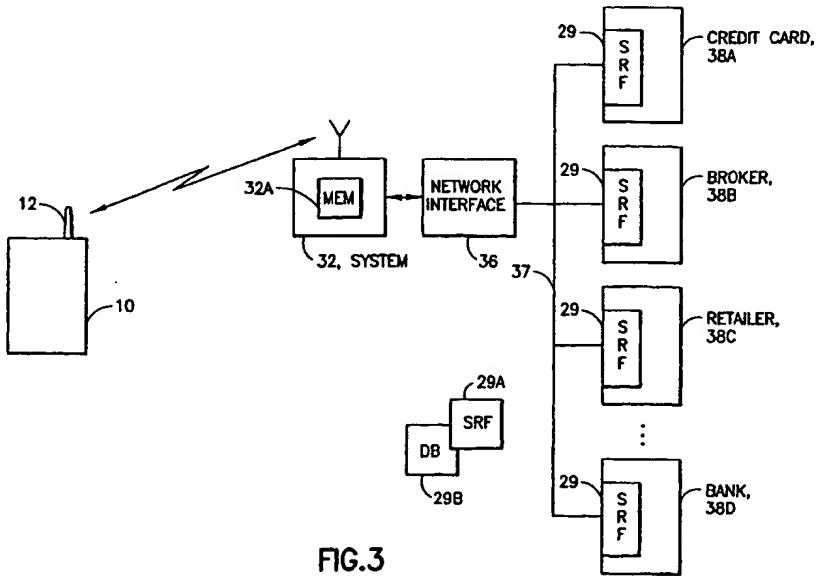

FIG.3

EP 1 074 974 A2

Description

[0001] This invention relates generally to biometric systems and methods and, in particular, to systems that identify a speaker by the automatic recognition of the speaker's voice and, more particularly, to a wireless telecommunications system employing voice recognition.

[0002] Biometric systems typically employ and measure some physical characteristic of a particular individual to uniquely identify that individual. The characteristic could be, by example, a fingerprint, a retinal pattern, or a voice pattern. The use of this latter characteristic is especially attractive for those systems that already include a microphone, such as telecommunications system, as no hardware expense may need to be incurred in order to implement the identification system. After having uniquely identified a speaker as being a particular, authorized individual, the system can then grant the speaker access to some location or to some resource. That is, this type of biometric system can be viewed as an electronic, voice actuated lock.

[0003] One problem that arises in many such systems is that the system is trained to recognize a particular speaker using a limited set of spoken words. For example, the speaker may be expected to say his or her name, and/or some predetermined password. While this approach may be suitable for many applications, in other applications the limited set of words used for identification may not be desirable, and may in fact lead some other persons to attempt to defeat the voice recognition-based biometric system. For example, a person attempting to defeat the system may simply surreptitiously tape record a person speaking the word or words that the biometric system expects to be spoken, and then play back the authorized person's speech to the voice input transducer of the biometric system.

It is well known in the mobile telecommunications art to provide a mobile telephone, such as a vehicle-installed cellular telephone, with a voice recognition capability in order to replace or augment the normal user input device(s). For example, the user can dial a number by speaking the digits, or by speaking a name having a stored telephone number. Some commands could be given to the telephone in the same manner.

[0004] In general, current user identification methods are based on measuring one static feature: e.g., a written password, a spoken password (voice recognition), a fingerprint, an image of the eye and so on. In the identifying situation the user knows what is measured and how.

[0005] It is an object of this invention to provide an improved biometric system, in particular a voice actuated recognition system, that relies on a random set of words and or images.

[0006] It is a further object of this invention to provide a mobile station having a speech transducer, and a method and apparatus to authenticate or authorize a user of a wireless telecommunication system to operate in, or through, or with a resource reachable through the wireless telecommunication system, only if the user's speech characteristics match pre-stored characteristics associated with word selected randomly from a training set of words.

[0007] The foregoing and other problems are overcome and the objects of the invention are realized by methods and apparatus in accordance with embodiments of this invention.

[0008] According to this invention, when a user enters an identifying situation he or she does not know beforehand what the identification stimulus will be and, thus, what the user's reaction or response will be. Using current technology a most straightforward way to implement the invention is with voice recognition. In this case the user is presented with a voice stimulus, or a text stimulus, or a graphical image stimulus, and the user reacts with his or her voice. The stimulus can be direct (e.g., the user speaks a displayed word) or indirect (e.g., the user responds to a question that only the user knows the answer to). Since even the correct user does not know beforehand the details of the identification situation, it becomes very difficult or impossible to know beforehand what the expected correct response will be.

[0009] A method is disclosed to authorize or authenticate a user of a wireless telecommunication system, and includes steps of (a) selecting a word at random from a set of reference words, or synthesizing a random reference word; (b) prompting the user to speak the reference word; and (c) authenticating the user to operate in or through or with a resource reachable through the wireless telecommunication system, only if the user's speech characteristics match predetermined characteristics associated with the reference word.

[0010] In one embodiment the steps of selecting or synthesizing, prompting and authenticating are performed in a mobile station having a speech transducer for inputting the user's speech, while in another embodiment at least one of the steps of selecting or synthesizing, prompting and authenticating are performed in a wireless telecommunications network that is coupled between the mobile station and a telephone network. In yet another embodiment at least one of the steps of selecting or synthesizing, prompting and authenticating are performed in a data communications network resource that is coupled through a data communications network, such as the Internet, and the wireless telecommunications network to the mobile station.

[0011] The step of prompting may include a step of displaying alphanumeric text and/or a graphical image to the user using a display of the mobile station.

[0012] The above set forth and other features of the invention are made more apparent in the ensuing Detailed Description of the Invention when read in conjunction with the attached Drawings, wherein:

Fig. 1 is a block diagram of a mobile station that is constructed and operated in accordance with this

invention;

Fig. 2 is an elevational view of the mobile station shown in Fig. 1, and which further illustrates a cellular communication system to which the mobile station is bidirectionally coupled through wireless RF links; and

Fig. 3 is block diagram that shows in greater detail a plurality of data communications network resources in accordance with further embodiments of this invention.

[0013] Reference is made to Figs. 1 and 2 for illustrating a wireless user terminal or mobile station 10, such as but not limited to a cellular radiotelephone or a personal communicator, that is suitable for practicing this invention. The mobile station 10 includes an antenna 12 for transmitting signals to and for receiving signals from a base site or base station 30. The base station 30 is a part of a wireless telecommunications network or system 32, that may include a mobile switching center (MSC) 34. The MSC 34 provides a connection to landline trunks, such as the public switched telephone network (PSTN) 35, when the mobile station 10 is involved in a call.

[0014] The mobile station includes a modulator (MOD) 14A, a transmitter 14, a receiver 16, a demodulator (DEMOD) 16A, and a controller 18 that provides signals to and receives signals from the transmitter 14 and receiver 16, respectively. These signals include signalling information in accordance with the air interface standard of the applicable cellular system, and also user speech and/or user generated data. The particular air interface standard and/or access type is not germane to the operation of this system, as mobile stations and wireless systems employing most if not all air interface standards and access types (e.g., TDMA, CDMA, FDMA, etc.) can benefit from the teachings of this invention.

[0015] It is understood that the controller 18 also includes the circuitry required for implementing the audio and logic functions of the mobile station. By example, the controller 18 may be comprised of a digital signal processor device, a microprocessor device, and various analog to digital converters, digital to analog converters, and other support circuits. The control and signal processing functions of the mobile station 10 are allocated between these devices according to their respective capabilities. In many embodiments the mobile station 10 will include a voice encoder/decoder (vocoder) 18A of any suitable type.

[0016] A user interface includes a conventional earphone or speaker 17, a conventional microphone 19, a display 20, and a user input device, typically a keypad 22, all of which are coupled to the controller 18. The keypad 22 includes the conventional numeric (0-9) and related keys (#,*) 22a, and other keys 22b used for operating the mobile station 10. These other keys 22b may include, by example, a SEND key, various menu scrolling and soft keys, and a PWR key. The mobile station 10 also includes a battery 26 for powering the various circuits that are required to operate the mobile station. The mobile station 10 also includes various memories, shown collectively as the memory 24, wherein are stored a plurality of constants and variables that are used by the controller 18 during the operation of the mobile station. The memory 24 may also store all or some of the values of various wireless system parameters and the number assignment module (NAM). An operating program for controlling the operation of controller 18 is also stored in the memory 24 (typically in a ROM device).

[0017] In accordance with the teachings of this invention, the controller 18 includes a speech recognition function (SRF) 29 that receives digitized input that originates from the microphone 19, and which is capable of processing the digitized input and for comparing the characteristics of the user's speech with pre-stored characteristics stored in the memory 24. If a match occurs then the controller 18 is operable to grant the speaker access to some resource, for example to a removable electronic card 28 which authorizes or enables the speaker to, in a typical application, make a telephone call from the mobile station 10. For example, the subscriber data required to make a telephone call, such as the Mobile Identification Number (MIN), and/or some authentication-related key or other data, can be stored in the card 28, and access to this information is only granted when the user speaks a word or words that are expected by the SRF 29, and which match predetermined enrollment (training) data already stored in the memory 24.

[0018] Further in accordance with this invention, the training data could as well be stored in some other memory, such as a memory 28A within the card 28, or in a memory 32A located in the system 32, or in some remote memory that is accessible through the system 32. For example, and referring specifically to Fig. 2, a memory 39 storing the training data set could be located in a data communications network (e.g., the Internet) entity or resource 38, which is accessible from the PSTN 35 through a network interface 36 (e.g., an Internet Service Provider or ISP), and a local area or wide area data communications network 37 (e.g., the Internet). In this case it can be appreciated that at least some of the data is packetized and sent in TCP/IP format.

[0019] In general, the identification system and software, as well as the prestored speech samples and characteristics may be located in the mobile station 10, in a server of the network 37 or the system 32, or in the system of a service provider.

[0020] In accordance with the an aspect of this invention the user can be prompted to speak one or a set of words, with the specific word to be spoken being selected randomly from the set of known words by the SRF

29. Assuming that the set of known words has a non-trivial number of elements, then it becomes difficult for another person to defeat the SRF 29 by recording a word or words expected to be spoken by the user.

[0021] The user can be prompted to speak the selected word or words in various ways. In a simplest way the SRF 29 displays the selected word on the display 20. Alternatively, the SRF 29 can use a speech synthesizer and the mobile station's speaker 17 to audibly prompt the user for the word to be spoken. In another embodiment the display 20 is used to present some graphical image corresponding to a word to be spoken (e.g., a tree). In a further embodiment some generic graphical image is used to suggest to the user a predetermined word to be spoken, and that was previously agreed upon during the training or enrollment stage. For example, it can be agreed upon that when presented with the graphical image of a tree the user will speak the word "birch", and that when presented with a graphical image of a city skyline the user will speak the word "Chicago". In this latter embodiment, and even if an unauthorized person where to gain possession of the user's mobile station 10, it is unlikely that the unauthorized person will give the correct reply word when presented with a particular graphical image or icon, let alone speak the reply word in a manner that would be recognized by the SRF 29 as a valid response.

[0022] If the set of training words are stored in the mobile station 10, whether in the memory 24 or the card 28, the words can be encrypted to prevent unauthorized access and/or modification.

[0023] Referring to Fig. 3, it can also be appreciated that the SRF 29 can be resident outside of the mobile station 10, such as at one or more network entities or resources 38A-38D (e.g., a credit card supplier, stock broker, retailer or bank.) In this embodiment, and assuming for example that the user wishes to access his account at the bank 38D, the SRF 29 signals back to the mobile station 10 a randomly selected word to be spoken by the user, via the network 37, network interface 36, and wireless system 32. The user speaks the word and, in one embodiment, the spectral and temporal characteristics of the user's utterance are transmitted from the mobile station 10 as a digital data stream (not as speech per se) to the SRF 29 of the bank 38D for processing and comparison. In another embodiment the user's spoken utterance is transmitted in a normal manner, such as by transmitting voice encoder/decoder (vocoder 18A) parameters, which are converted to speech in the system 32. This speech is then routed to the SRF 29 of the bank 38D for processing and comparison. It should be noted that the spectral and temporal and characteristics transmitted in the first embodiment could be the vocoder 18A output parameters as well, which are then transmitted on further to the SRF 29 of the band 38D, without being first converted to a speech signal in the system 32. In this case the necessary signalling protocol must first be defined and established so that the system 32 knows to bypass its speech decoder.

[0024] It is also within the scope of the teaching of this invention to provide a centralized SRF 29A, whose responsibility it is to authenticate users for other locations. For example, assume that the user of the mobile station 10 telephones the bank 38D and wishes to access an account. In this case the user authentication process is handled by the intervention of the SRF 29A which has a database (DB) 29B of recognition word sets and associated speech characteristics for a plurality of different users. The SRF 29A, after processing the user's speech signal, signals the bank 38D that the user is either authorized or is not authorized. This process could be handled in several ways, such as by connecting the user's call directly to the SRF 29A, or by forwarding the user's voice characteristics from the bank 38D to the SRF 29A. In either case the bank 38D is not required to have the SRF 29, nor are the other network resources 38A-38C.

[0025] It should be noted that the set of recognition words stored in the DB 29B could be different for every user. It should be further noted that this process implies that at some time the user interacts with the SRFs 29, or just with the SRF 29A, in order to execute an enrollment or training process whereby the user's database entries (set of recognition words and the associated speech temporal and spectral characteristics) are created. As was noted above, at least some of these speech characteristics could be based on or include voice encoder 18A parameters.

[0026] As an exemplary embodiment of this invention about 20-50 prestored voice samples can be used, and the stimulus and the sample are randomly or pseudorandomly selected among these (e.g., text-dependent speaker verification). In that the user records the samples himself or herself, the connection between the stimulus and the sample may be meaningful only for the user. Also, due to the provided stimulus the user is not required to memorize one or more passwords or numeric codes. Furthermore, there can be different sets of samples for different network services. For example, one set of samples may be used to obtain access to a network e-mail facility, while another set of samples may be used to obtain access to a network voice mail facility. As employed herein the term "random" is considered to encompass both truly random as well as pseudorandom.

[0027] For the case where speech synthesizing techniques improve sufficiently, it is also possible that the prestored samples are not required, but instead the system creates one or more synthesized reference word(s) that are compared to the user's voice response (text-independent speaker verification). The generated reference word is preferably generated randomly or pseudorandomly.

[0028] Furthermore, it should be appreciated that the teachings of this invention could be combined with the use of one or more other types of identification systems and techniques, such as fingerprint identification. Also, various ones of the stimulus types described above

could be used in combination. For example, the user may be presented with a randomly selected or generated alphanumeric string that the user is expected vocalize, as well as with a related or totally unrelated graphical image to which the user is expected to verbally respond.

[0029] While the invention has been described in the context of preferred and exemplary embodiments, it should be realized that a number of modifications to these teachings may occur to one skilled in the art. By example, any suitable speech processing techniques that are known for use in speech recognition systems can be employed, and the teachings of this invention are not limited for use to any specific technique.

[0030] Furthermore, while the user may be prompted to speak a reference "word", it can be appreciated that the "word" may actually be a phrase comprised of a plurality of words and also possibly numbers (e.g., a date, or an address).

[0031] Thus, while the invention has been particularly shown and described with respect to preferred embodiments thereof, it will be understood by those skilled in the art that changes in form and details may be made therein without departing from the scope and spirit of the invention.

Claims

1. A method to authenticate a user of a wireless telecommunication system, comprising steps of:

   selecting a word at random from a set of reference words;

   prompting the user to speak the selected word; and

   authenticating the user to operate in or through or with a resource reachable through the wireless telecommunication system, only if the user's speech characteristics match pre-stored characteristics associated with the selected word.

2. A method as in claim 1, wherein the steps of selecting, prompting and authenticating are performed in a mobile station having a speech transducer for inputting the user's speech.

3. A method as in claim 1, wherein at least one of the steps of selecting, prompting and authenticating are performed in a wireless telecommunications network coupled between a mobile station having a speech transducer for inputting the user's speech and a telephone network.

4. A method as in claim 1, wherein at least one of the steps of selecting, prompting and authenticating are performed in a data communications network resource that is coupled through a data communications network and a wireless telecommunications network to a mobile station having a speech transducer for inputting the user's speech.

5. A method as in claim 4, wherein the data communications network is comprised of the Internet.

6. A method as in claim 1, wherein the step of prompting includes a step of displaying alphanumeric text to the user using a display of a mobile station having a speech transducer for inputting the user's speech.

7. A method as in claim 1, wherein the step of prompting includes a step of displaying a graphical image to the user using a display of a mobile station having a speech transducer for inputting the user's speech.

8. A method to authenticate a user of a wireless telecommunication system, comprising steps of:

   synthesizing a random reference word;

   prompting the user to speak the synthesized reference word; and

   authenticating the user to operate in or through or with a resource reachable through the wireless telecommunication system, only if the user's speech characteristics match characteristics associated with the synthesized reference word.

9. A method as in claim 8, wherein the steps of synthesizing, prompting and authenticating are performed in a mobile station having a speech transducer for inputting the user's speech.

10. A method as in claim 8, wherein at least one of the steps of synthesizing, prompting and authenticating are performed in a wireless telecommunications network coupled between a mobile station having a speech transducer for inputting the user's speech and a telephone network.

11. A method as in claim 8, wherein at least one of the steps of synthesizing, prompting and authenticating are performed in a data communications network resource that is coupled through a data communications network and a wireless telecommunications network to a mobile station having a speech transducer for inputting the user's speech.

12. A method as in claim 11, wherein the data communications network is comprised of the Internet.

13. A method as in claim 8, wherein the step of prompting includes a step of displaying alphanumeric text to the user using a display of a mobile station having a speech transducer for inputting the user's speech.

14. A method as in claim 8, wherein the step of prompting includes a step of displaying a graphical image to the user using a display of a mobile station having a speech transducer for inputting the user's speech.

15. A wireless telecommunication system, comprising:

at least one base station;

a least one mobile station comprising a transceiver for conducting wireless communications with said base station, said mobile station further comprising a user interface and a microphone for inputting a users's speech;

a first subsystem coupled to said user interface for prompting the user to speak a reference word that is randomly selected from a set of reference words, or that is randomly generated; and

a second subsystem coupled to said microphone for authenticating the mobile station to operate in the wireless telecommunications system, or through the wireless telecommunications system, or with a resource that is reachable through the wireless telecommunication system, only if the user's speech characteristics match expected characteristics associated with the reference word.

16. A system as in claim 15, wherein one or both of the first and second subsystems are located in one of the mobile station, in the base station or in a controller coupled to the base station, or in a data communications network entity that is coupled through a data communications network to the wireless telecommunications system.

17. A system as in claim 16, wherein the data communications network is comprised of the Internet.

18. A system as in claim 15, wherein the first subsystem employs said user interface to at least one of present alphanumeric text to the user using a display of said mobile station, or to present a graphical image to the user using said display of said mobile station.

19. A system as in claim 15, wherein said mobile station further comprises a voice digitizing and processing system, and wherein said first subsystem further comprises means for transmitting an output of said voice digitizing and processing system to said base station, and wherein at least said second subsystem is located external to said mobile station.

20. A system as in claim 15, wherein at least said second subsystem is located in a network entity that is coupled to a data communications network that is bidirectionally coupled to said system.

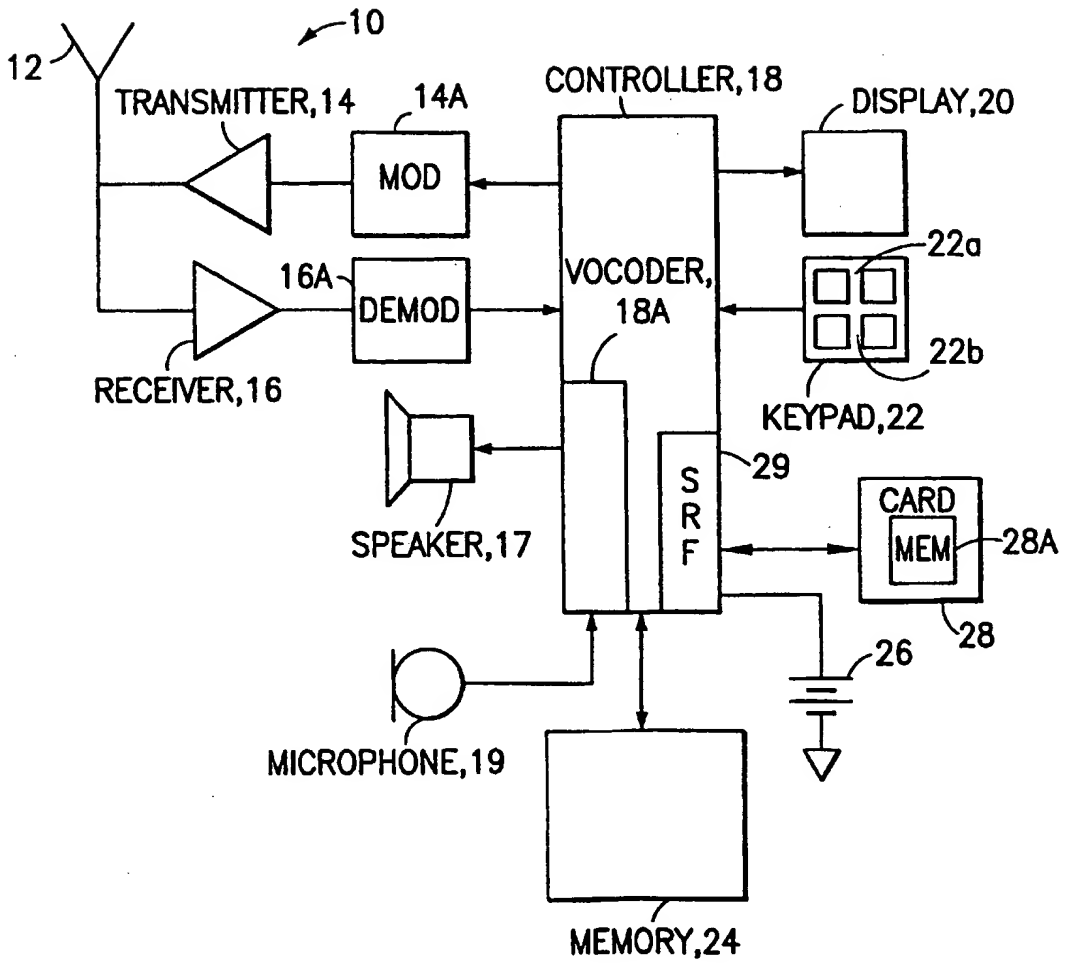
10
12
TRANSMITTER,14
14A
MOD
CONTROLLER,18
DISPLAY,20
16A
DEMOD
VOCODER,
18A
22a
22b
RECEIVER,16
KEYPAD,22
29
S
R
F
CARD
MEM
28A
28
SPEAKER,17
26
MICROPHONE,19
MEMORY,24

FIG. 1

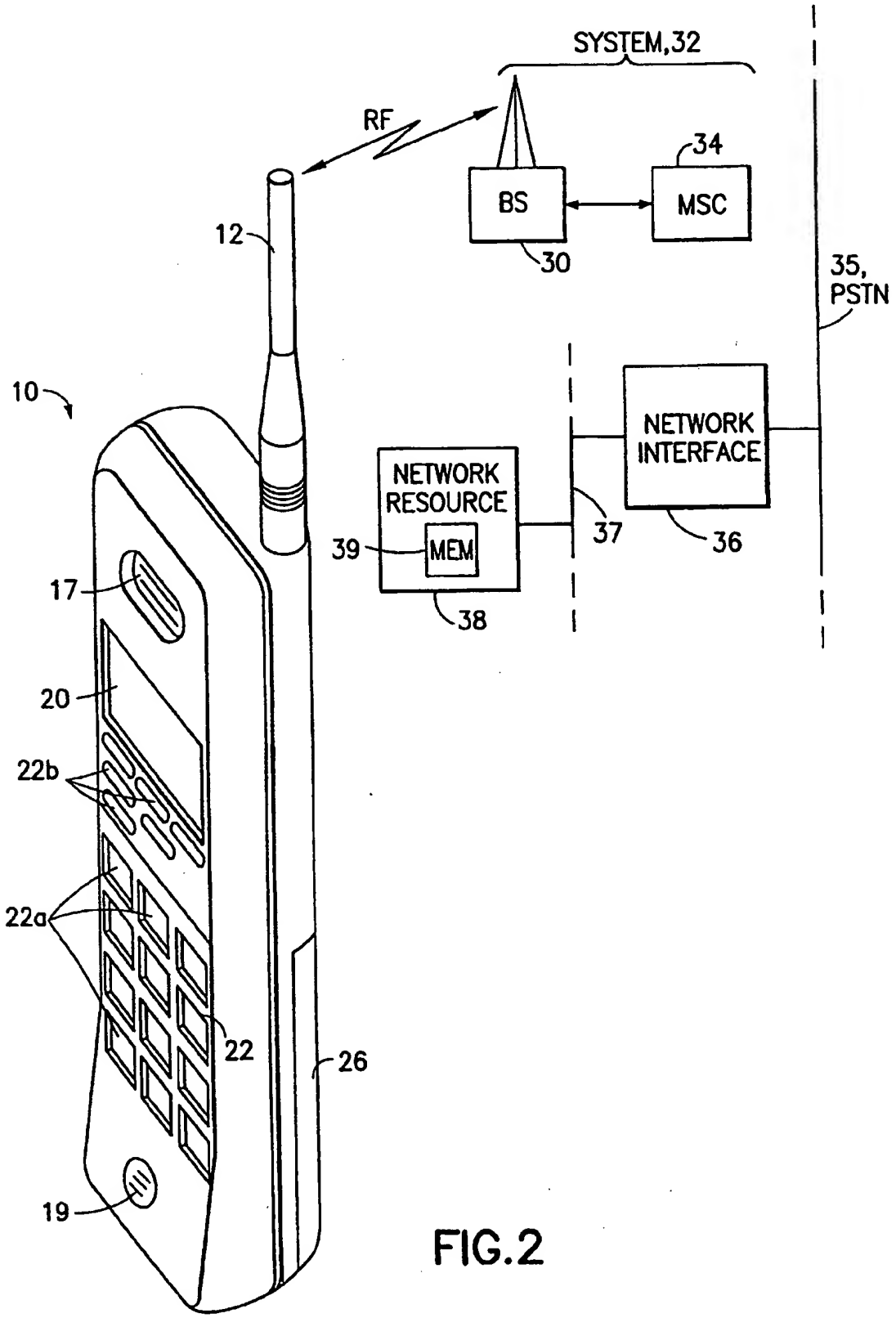
SYSTEM,32
RF
34
BS
MSC
12
30
35,
PSTN
10
NETWORK
INTERFACE
NETWORK
RESOURCE
MEM
39
37
36
38
17
20
22b
22a
22
26
19

FIG.2

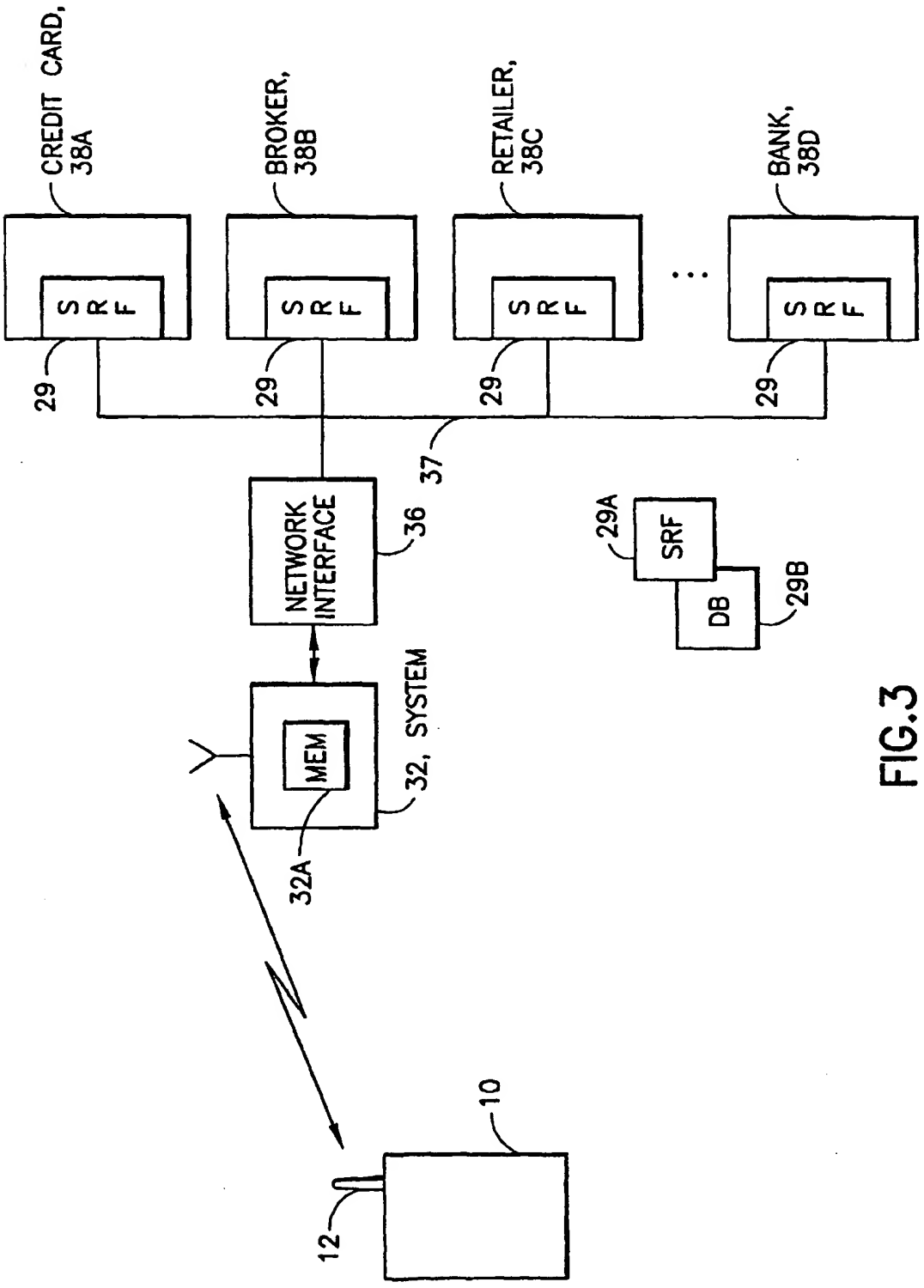
CREDIT CARD, 38A
BROKER, 38B
RETAILER, 38C
BANK, 38D
SRF
29
37
NETWORK INTERFACE
36
29A
SRF
DB
29B
MEM
32, SYSTEM
32A
10
12

FIG.3